

## PROJECT PROPOSAL

**Team members:** Varshitha Vasireddy, Vasu Dev, Pranav Vichare, Biswas Nandamuri

**Project title:** Disease prediction based on observed symptoms

**Description of the problem context:**

The main aim of this project is to predict disease with the symptoms said by the patients. Most of the diseases can be identified by the symptoms and identifying it correctly and treating it can save many lives. The top 3 diseases that match the symptoms can be predicted using machine learning.

**Type of the problem:** Supervised learning. Model is trained with the data of diseases and their symptoms. Methods including decision trees and knn are to be used.

**Initial thoughts on techniques that might be used:** Preprocessing the dataset to make it suitable for the model. Although we cannot use Imputation and Centering and Scaling techniques. Factor level processing can be used. PCA and LDA cannot be used as reducing the dimensions(symptoms) may affect the final predictions. Investigate the diseases and its symptoms and the number of diseases that the model can predict.

**Hyperlink of the problem (on kaggle):** <https://www.kaggle.com/datasets/itachi9604/disease-symptom-description-dataset?select=dataset.csv>